dnaMATE: a consensus melting temperature prediction server for short DNA sequences

Alejandro Panjkovich, Tomás Norambuena and Francisco Melo*

Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas,
Pontificia Universidad Católica de Chile, Alameda 340, Santiago, Chile

ABSTRACT

An accurate and robust large-scale melting temperature prediction server for short DNA sequences is dispatched. The server calculates a consensus melting temperature value using the nearest-neighbor model based on three independent thermodynamic data tables. The consensus method gives an accurate prediction of melting temperature, as it has been recently demonstrated in a benchmark performed using all available experimental data for DNA sequences within the length range of 16–30 nt. This constitutes the first web server that has been implemented to perform a large-scale calculation of melting temperatures in real time (up to 5000 DNA sequences can be submitted in a single run). The expected accuracy of calculations carried out by this server in the range of 50–600 mM monovalent salt concentration is that 89% of the melting temperature predictions will have an error or deviation of <5°C from experimental data. The server can be freely accessed at http://dna.bio.puc.cl/tm. html. The standalone executable versions of this software for LINUX, Macintosh and Windows platforms are also freely available at the same web site. Detailed further information supporting this server is available at the same web site referenced above.

INTRODUCTION

The accurate prediction of the DNA/DNA melting temperature is of paramount importance for the successful experimental implementation of several techniques in molecular biology that involve DNA/DNA hybridization, which include DNA microarrays, one locus or multiple loci PCR, quantitative PCR, DNA sequencing, Southern and northern blot, and any DNA hybridization-based technique. To date, different methods and several parameters have been described for the prediction of DNA/DNA melting temperatures. In a recent large-scale comparative assessment work, we have compared the melting temperature values obtained by different methods and/or parameterizations and demonstrated that significantly large differences, with a subsequent negative experimental outcome, could be obtained (1). Based on those results, we have derived a new consensus DNA/DNA melting temperature calculation method, which depends on different thermodynamic parameterizations and gives the most accurate melting temperature prediction values, according to an accuracy benchmark that was based on all available experimental values that involved hundreds of combinations of DNA sequences and salt concentrations (2–4).

In this paper, we dispatch a web server for the large-scale prediction of DNA melting temperatures, which is based on the consensus method described previously (1). The server was implemented to perform the simultaneous calculation of the melting temperatures for thousands of short DNA sequences in real time. It must be noted that this is the first web server available that provides this important feature.

METHODS

The consensus melting temperature predicted by the server is based on the nearest-neighbor thermodynamic calculations from three different experimentally derived thermodynamic data (5–7) and on the consensus map obtained from our large-scale comparative benchmark (1). A scheme showing how the melting temperature is calculated in conjunction with the consensus map is illustrated in Figure 1. The thermodynamic data used for the consensus melting temperature calculation include the tables from Breslauer (5), SantaLucia (6) and Sugimoto (7). The details of the consensus map derivation are available from our previous study (1) and also from the server web site http://dna.bio.puc.cl/tm.html.

The melting temperatures are calculated using the nearest-neighbor model and thermodynamic data as described previously (6). The equation that the server uses is as follows:

$$T_m = \frac{\sum (\Delta H_d) + \Delta H_i}{\sum (\Delta S_d) + \Delta S_i + \Delta S_{self} + R \times \ln(C_T/b)} + C_{Na^+}$$

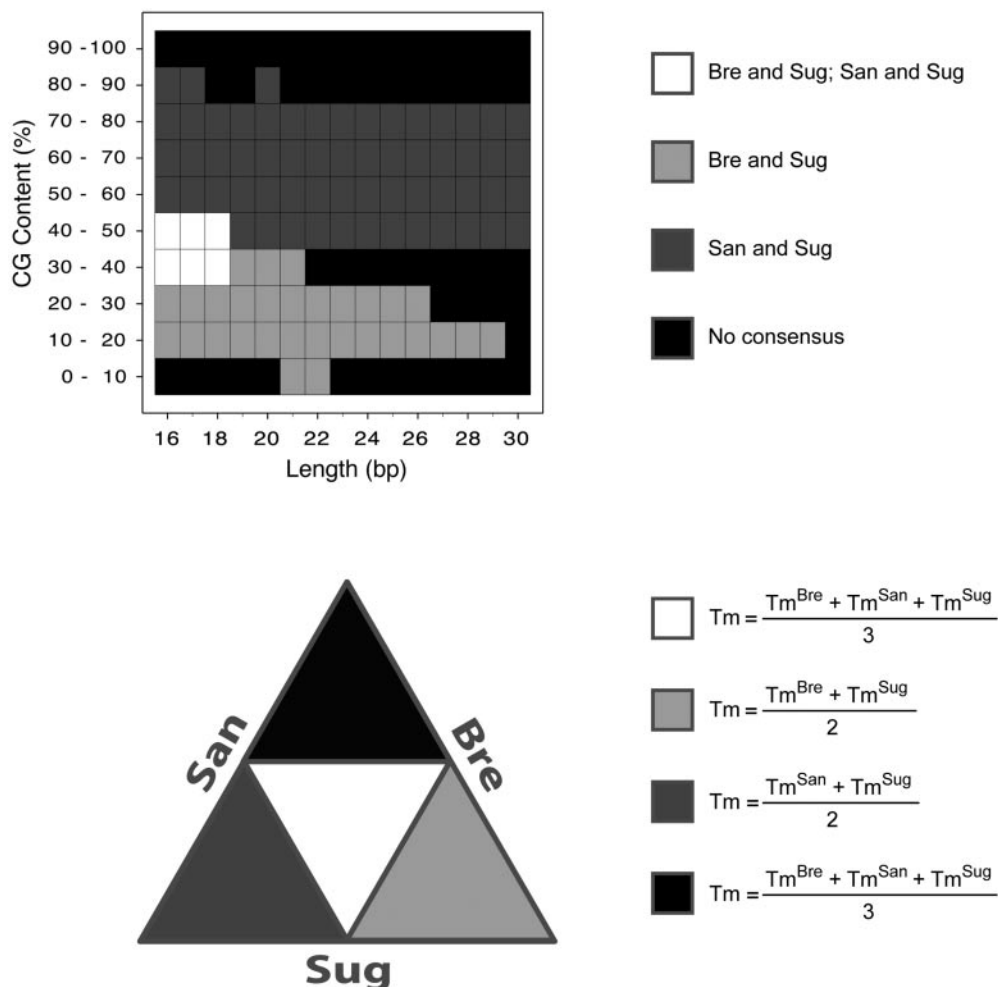*To whom correspondence should be addressed. Tel: +56 2 686 2279; Fax: +56 2 222 55 15; Email: fmelo@bio.puc.cl

Figure 1. Consensus T_m estimation method. Top panel: the consensus map from the previous comparative benchmark (1) is illustrated. In this benchmark, three thermodynamic data sets were compared: Bre stands for Breslauer (5); San stands for SantaLucia (9); and Sug stands for Sugimoto (7). In this map, four distinct regions were obtained: (i) simultaneously, Bre and Sug on the one hand, and San and Sug on the other, exhibited similar T_m values (white color); (ii) only Bre and Sug exhibited similar T_m values (light gray color); (iii) only San and Sug exhibited similar T_m values (dark gray color); and finally (iv) no consensus was observed among any of the methods (black color). Bre and San did not show a similar behavior in the complete range of sequence length and percentage of CG-content. Bottom panel: a graphical illustration of the different consensus map zones is shown. Each method is represented as a particular side of an equilateral triangle and the intersection among methods is shown with the corresponding color of the consensus map. The mathematical expressions used to calculate the consensus T_m at each zone are also indicated. In the case of San calculations, the most recent thermodynamic parameters (6) are being used by the server to calculate the consensus melting temperature. This modification with respect to our previous study (5) has further improved the accuracy of this server. The T_m estimations of oligonucleotides falling into the black regions of the consensus map by any of the methods could have a large error. The T_m estimation error at the other regions where some consensus was observed is expected to be small (below 3–5°C).

where sums of enthalpy (ΔH_d) and entropy (ΔS_d) are calculated over all internal nearest-neighbor doublets, ΔS_{self} is the entropic penalty for self-complementary sequences, and ΔH_i and ΔS_i are the sums of initiation enthalpies and entropies, respectively. R is the gas constant (fixed at 1.987 cal/K mol), C_T is the total strand concentration in molar units, $C_{\text{Na}+}$ is the salt adjustment factor and T_m is the melting temperature given in Kelvin units. Constant b adopts the value of 4 for non-self-complementary sequences or equal to 1 for duplexes of self-complementary strands or for duplexes when one of the strands is in significant excess. The Schildkraut–Lifson (8) equation is used as the salt adjustment factor, which corresponds to 16.6log [Na$^+$]. The thermodynamic calculations assume that the annealing occurs

in a buffered solution at pH near 7.0 and that a two-state transition occurs.

SERVER INPUT

The input of the server consists of a matrix containing one or more rows, each composed of three columns. The columns account for the following data: a particular DNA sequence (ranging between 16 and 30 nt), the DNA sequence concentration and the monovalent salt concentration, both in molar units. In the web server application, the total number of rows is limited to a maximum number of 5000 sequences. In the standalone version of the software, this limit only depends on the available memory of the computer where the software is

installed and executed (i.e. millions of DNA sequences for a typical personal computer).

SERVER OUTPUT

The output of the server consists of an equivalent row-sized matrix as the input, but containing several columns with calculated data, which include (i) the sequential number of the DNA sequence or sequential row number, (ii) the DNA sequence, (iii) the DNA sequence length, (iv) the oligo concentration, (v) the salt concentration, (vi) the CG-content of the DNA sequence, the melting temperature calculated using the thermodynamic data from (vii) Breslauer (5), (viii) SantaLucia (6), (ix) Sugimoto (7), (x) the consensus melting temperature calculated using a combination of the thermodynamic data that depends on the particular DNA sequence and the consensus map generated in our previous study (1), (xi) the consensus type that describes the experimental data used to calculate the consensus melting temperature, and (xii) a status message that reports the expected error of the T_m estimation. The user can choose if a simple or detailed output is provided as an HTML table. The server's web site contains online help support for each item.

IMPORTANT LIMITATIONS AND CONSIDERATIONS

Some important guidelines are recommended when using this server to obtain a high accuracy of melting temperature predictions. (i) Apply safely the current methods by considering the restrictions or limitations they have (i.e. avoid sequences that form stable alternative secondary structures, because such sequences are not going to follow a two-state transition, which is an important requirement of all methods that use the nearest-neighbor model to predict the melting temperature). (ii) Avoid using sequences that fall in those regions of oligonucleotide feature space where none of the current methods agrees (black regions of Figure 1A). (iii) If possible, use oligonucleotide sequences that fall in the middle range of CG-content and of a length 16–22mer (i.e. where most of the current melting temperature prediction methods agree). (iv) Salt correction is an important issue and the consensus map used by this melting temperature prediction method has been developed at low salt concentration, giving the best results when monovalent salt concentration is in the range 50–600 mM. Therefore, it is recommended to use this server in that salt concentration range to achieve a high accuracy or a low error in the melting temperature predictions. We are currently working in the derivation of a more sophisticated consensus map that not only takes into account the length and CG-content of the oligonucleotide and several thermodynamic tables but also the salt concentration and the salt adjustment factor. Therefore, there will be future improvements of the dnaMATE server.

REFERENCES

1. Panjkovich,A. and Melo,F. (2005) Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics*, **21**, 711–722.
2. Chiu,W.L.A.K., Sze,C.N., Ma,N.T., Chiu,L.F., Leung,C.W. and Au-Yeung,S.C.F. (2003) NTDB: thermodynamic database for nucleic acids, version 2.0. *Nucleic Acids Res.*, **31**, 483–485.
3. Owczarzy,R., Vallone,P.M., Gallo,F.J., Paner,T.M., Lane,M.J. and Benight,A.S. (1998) Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers*, **44**, 217–239.
4. Owczarzy,R., You,Y., Moreira,B.G., Manthey,J.A., Huang,L., Behlke,M.A. and Walder,J.A. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, **43**, 3537–3554.
5. Breslauer,K.J., Frank,R., Blocker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.
6. SantaLucia,J.J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
7. Sugimoto,N., Nakano,S., Yoneyama,M. and Honda,K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.*, **24**, 4501–4505.
8. Schildkraut,C. and Lifson,S. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*, **3**, 195–208.
9. SantaLucia,J.J., Allawi,H.T. and Seneviratne,P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.